ỨNG DỤNG KỸ THUẬT HỌC CHUYỂN GIAO CHO HỆ THỐNG PHÁT HIỆN TRỘM TRONG THỜI GIAN THỰC TRÊN THIẾT BỊ BIÊN CÓ TÀI NGUYÊN GIỚI HẠN

Lê Thị Trang [1*], Phan Thị Thể [2], Nguyễn Thị Vân Hảo[3]

[1] *Trường Đại học Công nghệ Đồng Nai*
[2] *Trường Đại học Sư phạm Kỹ thuật Tp. Hồ Chí Minh*
[3] *Trường Cao đẳng Kỹ nghệ 2*
*Tác giả liên hệ: Lê Thị Trang, *lethitrang@dntu.edu.vn*

THÔNG TIN CHUNG

Ngày nhận bài: 07/03/2024

Ngày nhận bài sửa: 16/04/2024

Ngày duyệt đăng: 07/05/2024

TỪ KHOÁ

Phát hiện kẻ trộm;
Điện toán biên;
Học chuyển giao;
Phát hiện đối tượng.

TÓM TẮT

Hiện nay, các camera giám sát ngày càng được sử dụng rộng rãi trong các căn hộ gia đình, do những tiến bộ công nghệ phần cứng, khả năng kết nối cao, chi phí thấp. Việc tích hợp các camera với thiết bị biên khác nhau vào hệ sinh thái nhà thông minh đang trở nên phổ biến nhằm tạo ra bộ điều khiển chung cho các thiết bị khác như: đèn, chuông cửa, nhiệt độ. Tuy nhiên, các camera do giới hạn về tài nguyên phần cứng nên thường chỉ hỗ trợ kỹ thuật nhận dạng chuyển động, người đơn giản. Trong nghiên cứu này, chúng tôi xây dựng mô hình nhận dạng trộm theo thời gian thực dựa trên kỹ thuật học chuyển giao (Transfer learning) trên thiết bị biên phổ biến Hệ thống đã được thử nghiệm trên các bộ dữ liệu mẫu và trên các mô hình YOLOv8n, EfficientDetD0, MobilenetV2 SSDLite cho thấy mô hình tốc độ độ xử lý thấp nhất là 3.6 FPS (Raspberry Pi 3 B+), 6.5 FPS (Raspberry Pi 4 B), 10.2 FPS (Jetson Nano) với độ chính xác AP (kiểm tra) trên 60%. Vì vậy, hệ thống khả thi khi triển khai vào thực tế.

1. GIỚI THIỆU

Ngày nay, sự tích hợp của AI với các hệ thống an ninh, giám sát đã tạo ra công cụ hữu ích trong việc quản lý tài sản của cá nhân và tổ chức. Các hệ thống giám sát thông minh thường sử dụng các mô hình trí tuệ nhân tạo để nhận dạng đối tượng như: trộm, cướp, người lạ, ... Ngoài ra, sự phát triển của các thiết bị biên (Edge devices), đặc biệt là trong bối cảnh Internet vạn vật (IoT), đã tạo ra xu hướng tích hợp các mô hình vào các thiết bị biên nhằm tăng cường khả năng xử lý tại chỗ không cần máy chủ, không sử dụng dịch vụ trên đám mây.

Thiết bị biên thường được trang bị một số phần cứng có khả năng xử lý dữ liệu tốt hơn các thiết bị điện tử truyền thống như CPU, GPU, RAM. Nhưng nhìn chung, tài nguyên trên thiết bị biên thường bị hạn chế so với máy tính, máy chủ. Vì vậy, các nghiên cứu về nhận dạng đối tượng hiện nay đang tập trung vào việc điều

chỉnh, giảm kích thước mô hình để có thể hoạt động theo thời gian thực trên các thiết bị biên vẫn đảm bảo độ chính xác chấp nhận được Khi sử dụng camera để phát hiện trộm, có một số khó khăn và thách thức là:

+ Góc quan sát hạn chế: vị trí lắp đặt camera ảnh hưởng đến hình ảnh thu được, camera thường đặt vị trí góc cao nên hình ảnh thu được sẽ khó thấy được gương mặt kẻ trộm.

+ Ánh sáng yếu: trong điều kiện ánh sáng yếu, ánh sáng không đều hoặc không có ánh sáng, camera có thể gặp khó khăn trong việc ghi lại hình ảnh rõ ràng của kẻ trộm.

+ Giả mạo và che đậy: kẻ trộm sử dụng nhiều cách để giả mạo hoặc che đậy danh tính như: đeo khẩu trang, mặc áo trùm đầu, đội nón, mặc áo đen, cúi đầu, ... làm cho việc nhận diện chúng trở nên khó khăn hơn.

+ Camera cần kết nối đến hệ thống tập trung có cấu hình phần cứng mạnh (thường trên đám mây) để xử lý theo thời gian thực với độ chính xác cao hạn chế báo động giả do nhầm lẫn chuyển động của động vật, côn trùng hoặc các yếu tố môi trường với chuyển động của con người.

Trong báo cáo này, chúng tôi đề xuất mô hình sử dụng kỹ thuật học chuyển giao (transfer learning) để tạo ra mô hình nhận dạng trộm với độ chính xác tương đối phù hợp sử dụng trên thiết bị biên Raspberry Pi 3 Model B+ (RPi 3B+), Raspberry Pi 4 Model B (RPi 4B), NVIDIA Jetson Nano Developer Kit B01 (Jetson Nano) theo thời gian thực (mô tả hình 1). Phương pháp thực hiện bao gồm các nội dung sau: (1) Nhận dạng chuyển động từ video và (2) Phân tích đối tượng chuyển động có phải là trộm hay không.

Bài báo được trình bày với Phần 1 là giới thiệu, Phần 2 cung cấp cơ sở và nghiên cứu liên quan về các phương pháp và mô hình nhận dạng trộm. Phần 3 mô tả kỹ thuật chính xây dựng hệ thống. Phần 4 thử nghiệm và đánh giá. Cuối cùng, phần 5 kết luận và các hướng nghiên cứu trong tương lai.

2. NHỮNG CÔNG TRÌNH NGHIÊN CỨU LIÊN QUAN

Các hệ thống phát hiện xâm nhập bất hợp pháp, trộm đã được phát triển từ lâu bằng cách sử dụng các thiết bị điện tử như cảm biến chuyển động, cảm biến nhiệt, cảm biến mở cửa, cảm biến rung, cảm biến hồng ngoại hoặc kết hợp các cảm biến (Saranu et al., 2018). Tuy nhiên, các hệ thống này có nhược điểm là dễ nhận dạng nhầm.

Các nghiên cứu gần đây thường sử dụng trí tuệ nhân tạo kết hợp với hệ thống camera an ninh nhằm tăng độ chính xác và dễ dàng tùy chỉnh cho phù hợp với môi trường hoạt động. Các mô hình đang được phát triển theo hướng điều chỉnh tham số, giảm kích thước, tăng tốc độ xử lý để có thể hoạt động theo thời gian thực dựa trên kỹ thuật học chuyển giao. Đây là một kỹ thuật trong học máy, trong đó mô hình đã được huấn luyện trước trên một tập dữ liệu lớn được chuyển giao và sử dụng để huấn luyện một mô hình mới cho một bài toán tương tự hoặc liên quan trên tập dữ liệu nhỏ.

Các mô hình nhận dạng trộm thường bao gồm các chức năng là: phân tích chuyển động của người, phát hiện khẩu trang (mặt nạ), phát hiện vũ khí (Arora et al., 2021; Nighrunkar et al., 2022). Agarwal (2021) đề xuất hệ thống sẽ phát hiện chuyển động với sự trợ giúp của mạng nơ-ron tích chập và thông báo cảnh báo khi có trộm (Agarwal et al., 2021). Giải pháp tự động phát hiện và nhận biết các tình huống nguy hiểm do Grega và cộng sự (2016) đề xuất là nhận dạng cầm dao, súng trên tay. Kết quả giải thuật nhân dạng cầm dao tương đối tốt trong điều kiện hình ảnh chất lượng thấp từ camera (Grega et al., 2016).

Nghiên cứu của Verma và Dhillon (2017) ứng dụng Faster R-CNN để phát hiện súng cầm tay cho kết quả độ chính xác 93% (Verma, Dhillon et al., 2017). Arora và cộng sự (2021) đề xuất hệ thống phát hiện và giám sát hành vi trộm cắp sử dụng học máy bằng cách kết hợp nhận dạng chuyển động, người mang khẩu trang (mặt nạ) và có mang vũ khí (Arora et al., 2021). Nighrunkar và cộng sự (2022) đưa ra phương án tối ưu hóa sử dụng cGAN và YOLO để phát hiện hành vi trộm cắp với độ chính xác cGAN 97.8%, YOLOv3 89.9% và YOLOv5 87.5% (Nighrunkar et al., 2022).

Bộ dữ liệu UCF Crime là một tập dữ liệu được sử dụng trong lĩnh vực nhận dạng hành vi và phát hiện tội phạm trong video. Được tạo ra bởi đại học California, Santa Cruz (UC Santa Cruz), bộ dữ liệu này chứa tổng cộng 128 giờ từ 1900 đoạn phim dài giám sát trong thế giới thực và không bị cắt, được ghi lại từ nhiều camera giám sát khác nhau, mô tả 13 hành vi bất thường liên quan đến tội phạm khác nhau: lạm dụng, bắt giữ, đốt phá, đánh nhau, tai nạn trên đường, trộm cắp, nổ, tấn công, cướp, bắn súng, trộm cắp, trộm cắp trong cửa hàng và phá hoại. Bộ dữ liệu DCSASS được xây dựng từ UCF Crime trong đó khắc phục một số nhược điểm như các video bất thường có chứa nhiều tình huống bình thường dẫn đến độ chính xác thấp (Sultani et al., 2018). Mỗi đoạn phim là phiên bản rút gọn chỉ dài khoảng 4-5 giây DCSASS có tổng cộng 16853 video, trong đó 9676 video được gắn nhãn là bình thường và 7177 là bất thường. Đối với hành vi trộm cắp có 2048 đoạn phim, 965 đoạn phim gắn nhãn bất thường, 1983 đoạn phim gắn nhãn là bình thường và được chia thành 64 nội dung bao gồm trộm đồ xe ô tô (41), xe máy (19), trộm đồ trong quán ăn (1), vào nhà trộm (1), cửa hàng (1), xe đạp (1).

Khảo sát đến trộm cắp tài sản ở Việt Nam trên báo của báo vnexpress thì nội dung có sự khác biệt với tập dữ liệu DCSASS như: hành vi trộm cắp chủ yếu không phải là trộm đồ xe ô tô mà thực hiện trong nhà, mang nón bảo hiểm, khẩu trang. Hình 1 mô tả trường hợp trộm xe máy ở nước ngoài cho thấy kẻ trộm không mang nón bảo hiểm, khẩu trang như ở trong nước. Do đó, có thể thấy bộ dữ liệu hình ảnh về trộm phục vụ cho việc xây dựng mô hình phụ thuộc nhiều vào môi trường quan sát, thu thập.

(a) (b)

Hình 1. Thể hiện hình ảnh khác nhau giữa trộm xe máy ở nước ngoài (a) trong tập dữ liệu DCSASS và ở trong nước (b) trên báo VnExpress.

3. PHƯƠNG PHÁP ĐỀ XUẤT

Trong phần này, chúng tôi mô tả chi tiết phương pháp bao gồm hai chức năng là nhận dạng đối tượng chuyển động trong video và sử dụng mô hình học chuyển giao để xác định trộm.

3.1 Nhận dạng đối tượng chuyển động trong video

Nhận dạng chuyển động, là bước cơ bản trong kỹ thuật giám sát bằng video, nhằm mục đích phát hiện di chuyển của các đối tượng hoặc các thay đổi trong khung hình xác định. Các khó khăn gặp phải với phương pháp phát hiện chuyển động là nhiễu nguồn, hình nền phức tạp, sự thay đổi trong ánh sáng của cảnh, bóng của vật thể tỉnh và vật thể chuyển động.

Nhiều nghiên cứu về phát hiện chuyển động đã được thực hiện trong thập kỷ qua, các phương pháp được phân thành ba loại chính là background subtraction (loại bỏ nền cố định từ các khung hình để tìm ra các vùng chuyển động trong video); frame difference (phát hiện chuyển động trong video bằng cách so sánh sự

khác biệt giữa các khung hình gần nhau); Optical flow (theo dõi chuyển động của điểm pixel qua các khung hình liên tiếp để xác định hướng và tốc độ của sự di chuyển).

Theo nghiên cứu của Sehairi (2017) kỹ thuật frame difference có tốc độ xử lý các khung hình cao (Fps) và khả năng phát hiện chuyển động tốt trong điều kiện ánh sáng yếu hoặc hoàn toàn thiếu ánh sáng tự nhiên (night video) (Sehairi et al., 2017). Do đó, kỹ thuật temporal differencing phù hợp khi sử dụng với các thiết bị biên với phần cứng giới hạn và phù hợp với môi trường ánh sáng ban đêm là thời điểm kẻ trộm thường thực hiện.

Thuật toán frame difference so sánh hai khung hình liên tiếp trong video và xác định các pixel đã thay đổi. Các pixel đã thay đổi được cho là đại diện cho các đối tượng đang chuyển động. Công thức toán học (1) tính độ sai khác Δ_t giữa hai pixel $P_t(x, y)$ trong khung hình hiện tại và $P_{t-1}(x, y)$ khung hình trước đó là:

$$\Delta_t = |P_t(x, y) - P_{t-1}(x, y)| \quad (1)$$

Trong đó:

$P_t(x, y)$ là giá trị pixel tại vị trí (x, y) trong khung hình hiện tại.

$P_{t-1}(x, y)$ là giá trị pixel tương ứng tại vị trí (x, y) trong khung hình liền trước.

Sự chuyển động của đối tượng M_t được tính bằng cách so sánh giá trị của Δ_t với ngưỡng ∂ cho trước.

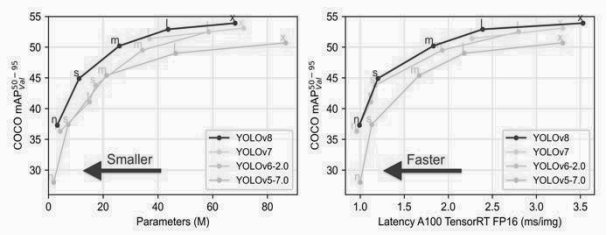
$$M_t = \begin{cases} 1, \Delta_t \geq \partial \\ 0, ngược\ lại \end{cases} \quad (2)$$

Do đối tượng phát hiện chuyển động của hệ thống là kẻ trộm cũng tương đồng với người đi bộ nên ngưỡng ∂=30 là phù hợp theo nghiên cứu của (Barbu et al., 2014).

3.2 Xây dựng mô hình học chuyển giao

Theo nghiên cứu của Kamath và Renuka, (2023) các mô hình nhận dạng đối tượng hiệu quả trên thiết bị biên là YOLOv7-X, EfficientDet-D0, MobilenetV2-SSDLite. Các mô hình trên được xây dựng trên bộ dữ liệu MS-COCO, trong đó lớp người có số lượng ảnh huấn luyện lớn nhất là 262465 so với các lớp còn lại (Aslam, Curry et al., 2021) và điều này phù hợp với ngữ cảnh của nghiên cứu là nhận dạng kẻ trộm. Các mô hình thực hiện thử nghiệm như sau:

- Yolov8n: YOLO (You Only Look Once) là một trong những mô hình mạnh mẽ để phát hiện đối tượng nhanh chóng và chính xác trong lĩnh vực thị giác máy tính. YOLO là một thuật toán One State Dectector, nghĩa là dự đoán nhãn và vị trí của đối tượng trong toàn bộ bức ảnh chỉ với một lần chạy thuật toán duy nhất, điều này giúp cho thời gian xử lý của YOLO rất nhanh, phù hợp với các ứng dụng cần chạy thời gian thực. Phiên bản YOLOv8 có nhiều cải tiến về kiến trúc giúp nó có tốc độ nhận dạng và độ chính xác cao hơn so với các phiên bản trước. Hình 2 mô tả mối quan hệ giữa số lượng tham số (params), thời gian xử lý trên mỗi hình ảnh (ms/img) và mAPval50-95 (là một chỉ số đo lường độ chính xác của các mô hình phát hiện đối tượng, được tính toán bằng cách lấy trung bình độ chính xác trung bình trên một loạt các ngưỡng giao thoa trên tỉ lệ giữa diện tích giao nhau và hợp nhất -IoU từ 0.5 đến 0.95 với bước tăng 0,05) cho bốn thuật toán YOLOv5, YOLO6, YOLOv7 và YOLOv8. Kết quả hình 1 cho thấy, YOLOv8 vượt qua các thuật toán khác về chỉ số mAPval50-95 trong cùng lượng tham số và thời gian. Yolov8n là phiên bản nhỏ gọn của mô hình Yolov8, được tối ưu hóa cho các thiết bị có hiệu năng thấp như điện thoại thông minh và thiết bị biên. Nó vẫn giữ được khả năng phát hiện đối tượng chính xác tốt nhưng với tốc độ xử lý nhanh và tiêu thụ ít năng lượng.



Hình 2. So sánh hiệu xuất hoạt động các phiên bản YOLO khác nhau nguồn Ultralytics: NEW - YOLOv8 (2024)

- MobilenetV2-SSDLite: là một mô hình phát hiện đối tượng được xây dựng dựa trên sự kết hợp giữa hai thành phần chính là MobileNetV2 và SSDLite. Mô hình thường được sử dụng trong các ứng dụng yêu cầu tính toán nhanh và tốc độ nhận dạng cao, hiệu quả để phát hiện đối tượng trên các thiết bị có tài nguyên hạn chế như điện thoại di động hoặc thiết bị biên (Sandler et al., 2018)

- EfficientDet-D0: EfficientDet là một thuật toán phát hiện đối tượng được phát triển bởi Google AI vào năm 2020. Nó được thiết kế để đạt được hiệu quả cao về cả tốc độ và độ chính xác. EfficientDet-D0 là phiên bản nhỏ nhất trong loạt mô hình EfficientDet, thường được sử dụng khi cần một mô hình nhẹ nhàng với tốc độ dự đoán nhanh và yêu cầu bộ nhớ ít. Mặc dù có ít tham số hơn so với các phiên bản lớn hơn, nhưng EfficientDet-D0 vẫn có khả năng phát hiện đối tượng hiệu quả trên nhiều tác vụ và tập dữ liệu khác nhau.

Mô hình YOLOv8n, EfficientDet-D0, và MobilenetV2-SSDLite là mô hình cơ sở để thực hiện học chuyển giao với bộ dữ liệu trộm trên các thiết bị biên để đánh giá tốc độ xử lý và độ chính xác.

4. THỬ NGHIỆM VÀ ĐÁNH GIÁ

4.1 Thử nghiệm

Bộ dữ liệu thử nghiệm gồm 598 ảnh được lấy ra từ các đoạn phim liên quan đến trộm cắp tài sản của báo vnexpress (n.d.) với môi trường là trộm trong nhà, ánh sáng tối, hình dạng trộm thường mặc áo khoác, đội nón, đeo khẩu trang, một vài trường hợp cầm dao (hình 3).

(a) (b)

Hình 3. Ảnh được trích xuất từ đoạn phim dùng cho quá trình huấn luyện, (a) là không có trộm, (b) có trộm

Bộ dữ liệu được chia thành ba tập dữ liệu con là: tập huấn luyện 418 ảnh (70%); tập tập xác thực 121 ảnh (20%); tập kiểm tra 59 ảnh (10%). Các mô hình thử nghiệm được được thực hiện trên máy tính PC cấu hình (CPU G3240 @ 3.10GHz, GPU GTX 1060 3G, 12GB Ram, 1TB HDD). Các thiết bị biên được sử dụng để thử nghiệm cho các mô hình sau khi học chuyển giao là RPi 3B+, RPi 4B, Jetson Nano. Đây là các thiết bị phổ biến trên thị trường và giá thành không cao. Chi tiết thông tin cấu hình ở bảng 1.

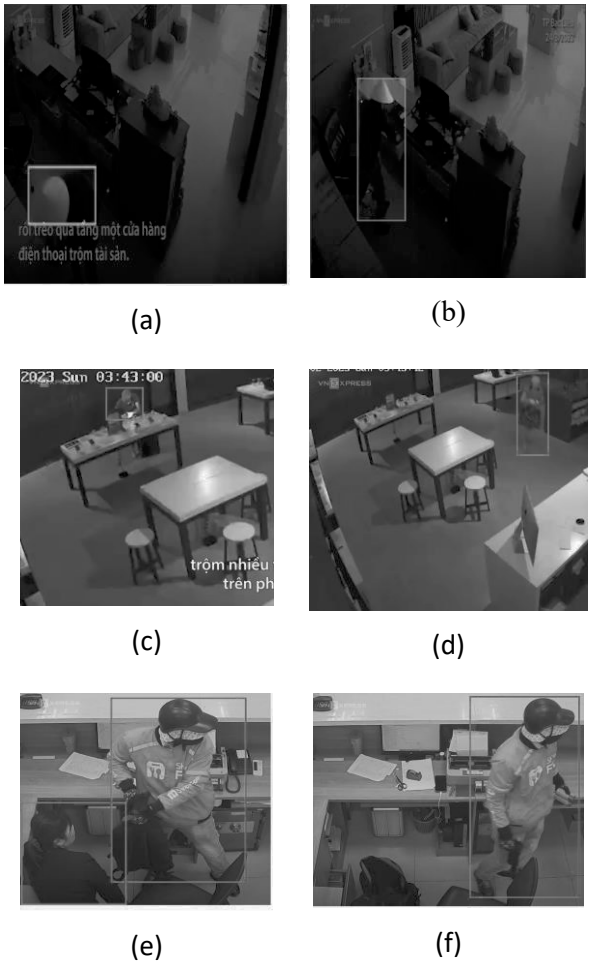
Bảng 1. Danh sách cấu hình các thiết bị

	RPi 3B+	**RPi 4B**	**Jetson Nano**
CPU	Quad core Cortex A53 1.4GHz	Quad core Cortex A72 1.5GHz	Quad-core ARM A57 1.43 GHz
GPU	Không	Không	128-core Maxwell
RAM	4GB	4GB	4GB
HDD	32GB	32GB	32GB

Hiệu quả của mô hình được đo lường bằng chỉ số AP (Average Precision) và tốc độ xử lý các khung hình thiết bị trên giây (Fps) của thiết bị.

4.2 Đánh giá kết quả

Theo phân tích kết quả bảng 2, các mô hình sao khi học chuyển giao độ chính xác AP tại tập huấn luyện tương đối trên 80%, tập kiểm thử thấp hơn chỉ trên 76%. Độ đo F1-Score cho thấy mô hình được học chuyển giao từ YOLOv8n có tốt hơn các mô hình còn lại. Các trường hợp nhận dạng sai thường do hình ảnh mờ, không rõ, kẻ trộm đứng gần đối tượng khác (tủ, bàn, ghế) hay nhầm lẫn trộm và người có hành động tương tự trong ảnh (hình 4)



(a) (b) (c) (d) (e) (f)

Hình 4. Ảnh nhận dạng sai (a)(c)(e) và ảnh nhận dạng đúng (b)(d)(f) trong cùng đoạn phim

Bảng 2. Kết quả độ đo chính xác của các mô hình

Thuật toán	AP train (%)	AP test (%)	F1 Score (%)
YOLOv8n	94.2	92.1	90.9
EfficientDet-D0	87.3	84.6	82.4
MobilenetV2 SSDLite	80.4	76.8	72.6

Đối với tốc độ xử lý (FPS) trên các thiết bị (bảng 3), mô hình YOLOv8n có tốc độ cao hơn các mô hình khác từ 1-2 Fps (Rpi 3B+), 3-4 Fps (RPi 4B), 5-7 Fps (Jetson Nano).

Bảng 3. Kết quả độ đo tốc độ xử lý của các mô hình

	Rpi 3B+ (FPS)	RPi 4B (FPS)	Jetson Nano (FPS)
YOLOv8n	5.3	9.5	15.4
EfficientDet DetD0	3.6	6.5	10.2
MobilenetV2 SSDLite	4.4	5.2	8.5

5. KẾT LUẬN

5.1 Kết luận

Mặc dù kỹ thuật nhận dạng đối tượng trong ảnh, video đã có sự phát triển mạnh mẽ trong hơn trong thập kỷ qua, nhưng nhiều mô hình deep learning có cấu trúc phức tạp, yêu cầu nhiều tài nguyên tính toán để huấn luyện và triển khai, đặc biệt là trên các thiết bị có tài nguyên hạn chế như thiết bị di động hoặc thiết bị biên. Trong nghiên cứu này, chúng tôi đề xuất xây dựng hệ thống phát hiện trộm thời gian thực trên thiết bị biên sử dụng kỹ thuật học chuyển giao với dữ liệu là kẻ trộm trên các mô

hình YOLOv8n, EfficientDetD0, MobilenetV2 SSDLite. Kết quả cho thấy mô hình YOLOv8n có độ chính xác tốt và thời gian xử lý 9.5 FPS trên thiết bị Raspberry Pi 4 Model B. Đây là thiết bị đang phổ biến trên thị trường với chi phí thấp. Vì vậy, nghiên cứu khả thi khi áp dụng để xây dựng hệ thống phát hiện trộm trên thực tế.

5.2 Hướng phát triển

Bộ dữ liệu nghiên cứu được lựa chọn từ đoạn phim quay lại cảnh trộm cấp chủ yếu trong nhà nên cần phải thu thập thêm các dữ liệu về các môi trường nhận dạng trộm khác nhau như: trộm ở hàng rào, trộm xe máy, ... Ngoài ra, kết quả nghiên cứu cho thấy có thể sử dụng kỹ thuật tương tự để xây dựng hệ thống nhận dạng đối tượng khác trên thiết bị biên.

TÀI LIỆU THAM KHẢO

Agarwal, V. (2021, August 3). Identity Theft Detection Using Machine Learning. *International Journal for Research in Applied Science and Engineering Technology*, 9(8), 1943–1946. https://doi.org/10.22214/ijraset.2021.37696

Anggraini, N., Ramadhani, S. H., Wardhani, L. K., Hakiem, N., Shofi, I. M., & Rosyadi, M. T. (2022, September 13). Development of Face Mask Detection using SSDLite MobilenetV3 Small on Raspberry Pi 4. 2022 5th International *Conference of Computer and Informatics Engineering* (IC2IE). https://doi.org/10.1109/ic2ie56416.2022.9970078

Arora, J., Bangroo, A., & Garg, S. (2021, December 1). Theft Detection and Monitoring System Using Machine Learning. Emerging Research in Computing, *Information, Communication and Applications*, 957–966. https://doi.org/10.1007/978-981-16-1342-5_76

Aslam, A., & Curry, E. (2021, February). A Survey on Object Detection for the Internet of Multimedia Things (IoMT) using Deep Learning and Event-based Middleware: Approaches, Challenges, and Future Directions. *Image and Vision Computing*, 106, 104095. https://doi.org/10.1016/j.imavis.2020.104095

Barbu, T. (2014, May). Pedestrian detection and tracking using temporal differencing and HOG features. *Computers & Electrical Engineering*, 40(4), 1072–1079. https://doi.org/10.1016/j.compeleceng.2013.12.004

Grega, M., Matiolański, A., Guzik, P., & Leszczuk, M. (2016, January 1). *Automated Detection of Firearms and Knives in a CCTV Image*. Sensors, 16(1), 47. https://doi.org/10.3390/s16010047

Kamath, V., & Renuka, A. (2023, April). Deep learning based object detection for resource constrained devices: Systematic review, future trends and challenges ahead. *Neurocomputing,* 531, 34–60. https://doi.org/10.1016/j.neucom.2023.02.006

Nighrunkar, M., Mahajan, S., Kulkarni, A., & Joshi, A. (2022). Theft Detection: An Optimized Approach Using cGAN and YOLO. *Advancements in Interdisciplinary Research*, 325–332. https://doi.org/10.1007/978-3-031-23724-9_30

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018, June). MobileNetV2: *Inverted Residuals and Linear Bottlenecks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognitio*n. https://doi.org/10.1109/cvpr.2018.00474

Saranu, P. N., Abirami, G., Sivakumar, S., Ramesh, K. M., Arul, U., & Seetha, J. (2018, February). Theft Detection System using PIR Sensor. 2018 4th *International Conference on Electrical Energy Systems* (ICEES).

https://doi.org/10.1109/icees.2018.8443215

Sehairi, K., Chouireb, F., & Meunier, J. (2017, April 25). Comparative study of motion detection methods for video surveillance systems. *Journal of Electronic Imaging*, 26(2), 023025. https://doi.org/10.1117/1.jei.26.2.023025

Sultani, W., Chen, C., & Shah, M. (2018, June). Real-World Anomaly Detection in Surveillance Videos. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/cvpr.2018.00678

U. (2024, January 10). Ultralytics/ultralytics: NEW - YOLOv8 in PyTorch > ONNX > OpenVINO > CoreML > TFLite. GitHub. https://github.com/ultralytics/ultralytics

V. (n.d.). Tin tức Trộm cắp tài sản mới nhất hôm nay trên VnExpress. vnexpress.net. https://vnexpress.net/chu-de/trom-cap-tai-san-5399

Verma, G. K., & Dhillon, A. (2017, November 24). A Handheld Gun Detection using Faster R-CNN Deep Learning. Proceedings of the 7th International Conference on Computer and Communication Technology. https://doi.org/10.1145/3154979.31

BUILDING A REAL-TIME THEFT DETECTION SYSTEM FOR EDGE DEVICES WITH LIMITED RESOURCES

Le Thi Trang [1*], Phan Thi The [2], Nguyen Thi Van Hao[3]

[1] *Dong Nai Technology University*
[2] *Ho Chi Minh City University of Technology and Education*
[3] *College of Technology II*
* Corresponding author: Le Thi Trang, *lethitrang@dntu.edu.vn*

GENERAL INFORMATION

Received date: 07/03/2024

Revised date: 16/04/2024

Accepted date: 07/05/2024

KEYWORD

Thief detection;
Edge Computing;
Transfer learning;
Object detection.

ABSTRACT

Currently, surveillance cameras are increasingly widely used in family apartments, due to advances in hardware technology, high connectivity, and low cost. Integrating cameras with different edge devices into the smart home ecosystem is becoming popular to create a common controller for other devices such as lights, doorbells, and temperature. However, cameras, due to limited hardware resources, usually only support simple motion and person recognition techniques. In this study, we build a real-time thief detection model based on transfer learning techniques on popular edge devices. Experimental results show that the model has the lowest inference speed 3.6 FPS (Raspberry Pi 3 B+), 6.5 FPS (Raspberry Pi 4 B), and 10.2 FPS (Jetson Nano), with AP (test) accuracy over 60%. Therefore, the system is feasible when deployed in practice.